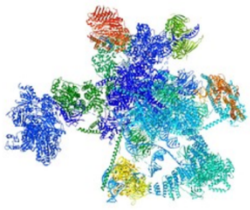
Virus: Norovirus
https://www.ncbi.nlm.nih.gov/protein/AKN44263
Protein target: VPg

```
>YP_009701442.1 VPg [Norovirus GI]
GKNKGKTKKGRGRKSNFNAFSRRGLSDEEYEEYKKIREEKSGNYSIQEYLEDRQRYEEELAEVQAGGDGG
IGETEAEIRHRVFYKSKSGMRKQRQEERRQLGLVSGSEIRKRKPIDWTPPKNDWSEDTRTVNYDEHISFE
```

For this discussion, I searched through the NCBI taxonomy browser for a virus containing a small protein of less than 125 residues in length. I found a VPg protein sequence from the Norovirus and I decided to run my analysis on it. Running the FASTA through PDB-Blast to see if there are any similar sequences in non-redundant databases and I found that the highest SeqID I found was 23%. Only the top match is shown below.

PDB-Blast Search



5XJC: Entity 38 containing Chain v Download File | View File

Cryo-EM structure of the human spliceosome just prior to exon ligation at 3.6 angstrom

Zhang, X., Yan, C., Hang, J., Finci, L.I., Lei, J., Shi, Y.

(2017) Cell **169** 918-929.e14

Released: 7/5/2017
Method: Electron Microscopy
Resolution: 3.6 Å
Residue Count: 24811

Macromolecule:
Unique protein chains: 36
Unique nucleic acid chains: 3
Unique Ligands: ADP, ATP, GTP, I6P, MG, SEP, UNK, ZN

Length: 100 **E-value:** 0.244989 **Score:** 32.7278bits (73) **Identities:** 23/100 (23%) **Positives:** 47/100 (47%) **Gaps:** 3/100 (3%)

```
      41        50        60        70        80        90        100       110       120       130       13
      |   .   |   .   |   .   |   .   |   .   |   .   |   .   |   .   |   .   |   .   |   .   |
Query SGNYSIQEYLEDRQRYEEELAEVQAGGDGGI-GETEAEIRHRVFYKSKSGMRKQRQEERRQLGLVSGSEIRKRKPIDWTPPKNDWSEDTRTVNYDEHISF
      + ++ ++ Y+  + ++  E  E +   DG +     +  ++ V  + ++ + K   EE +++  +  SEI K     W PP     ++   V  DEHISF
Sbjct ASDFYLRYYVGHKGKFGHEFLEFEFRPDGKLRYANNSNYKNDVMIRKEAYVHKSVMEELKRI--IDDSEITKEDDALWPPPDRVGRQELEIVIGDEHISF
      |.    |   .   |   .   |   .   |   .   |   .   |   .   |   .   |   .   |   .   |   .   |   .   .|
      4     10        20        30        40        50        60        70        80        90        10
```

My next step was to run my sequence in FFAS03 and Phyre2 for fold recognition.

Phyre2

Top model

Top model

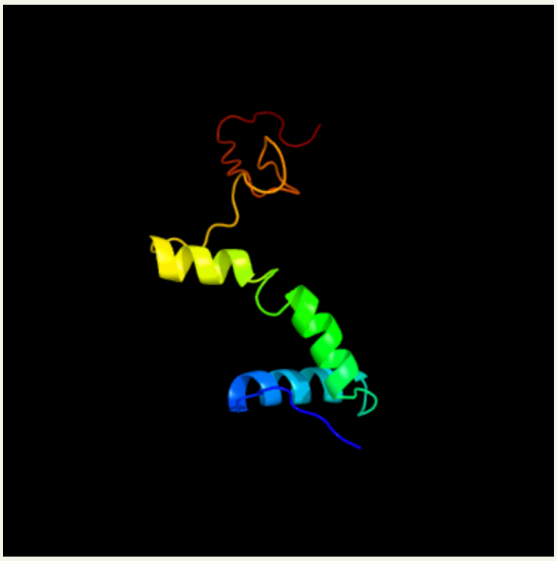


Image coloured by rainbow N → C terminus

Model dimensions (Å): **X**:66.763 **Y**:37.889 **Z**:35.244

Model (left) based on template c2m4gA_

Top template information

PDB header:viral protein
Chain: A: **PDB Molecule:**murine norovirus vpg protein;
PDBTitle: solution structure of the core domain (11-85) of the murine norovirus2 vpg protein

Confidence and coverage

Confidence: **100.0%** Coverage: **51%**

72 residues (51% of your sequence) have been modelled with 100.0% confidence by the single highest scoring template.

Phyre alarm

You may wish to submit your sequence to Phyrealarm. This will automatically scan your sequence every week for new potential templates as they appear in the Phyre2 library.

Warning: 62% of your sequence is predicted disordered. Disordered regions cannot be meaningfully predicted.

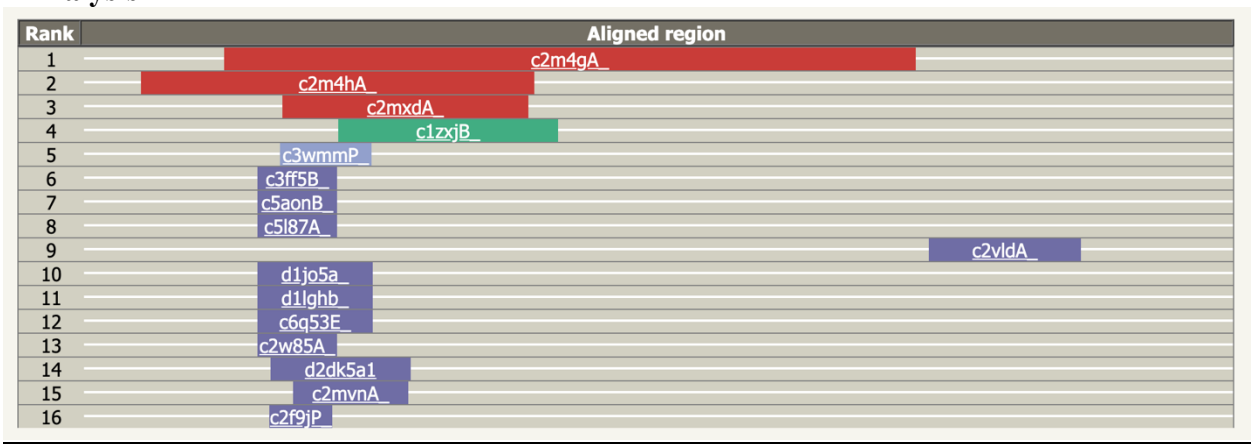
3D viewing

Interactive 3D view in JSmol

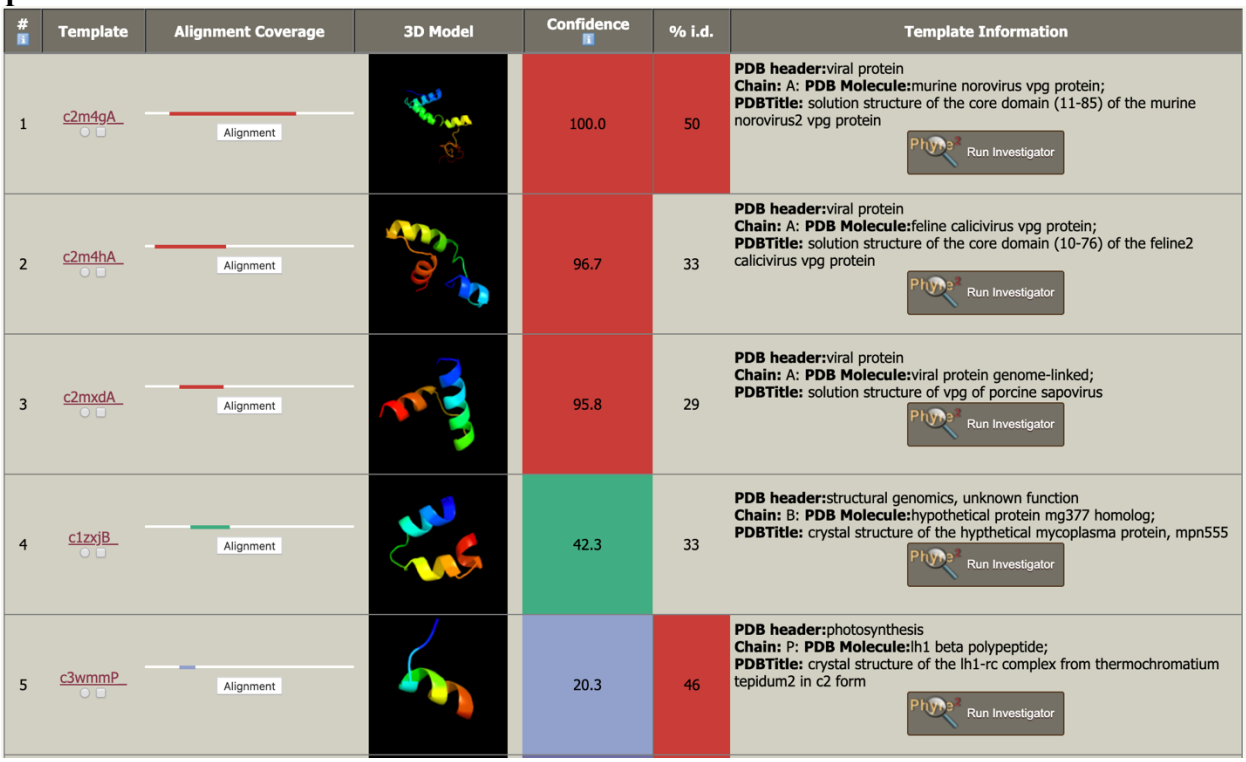
For other options to view your downloaded structure offline see the FAQ

Secondary structure/disorder prediction

Domain Analysis

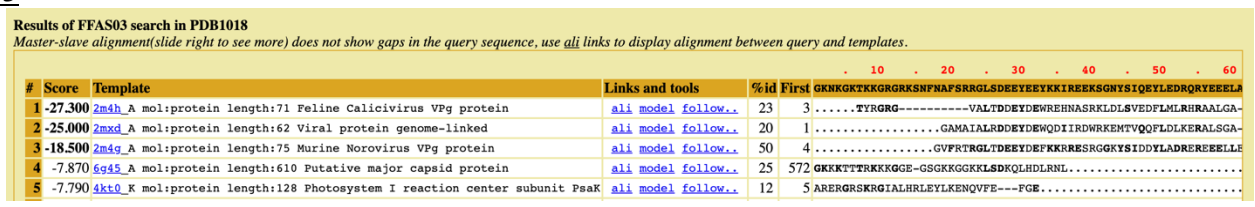


All template/model information



#	Template	Alignment Coverage	3D Model	Confidence	% i.d.	Template Information
1	c2m4gA_	Alignment		100.0	50	**PDB header:**viral protein **Chain:** A: **PDB Molecule:**murine norovirus vpg protein; **PDBTitle:** solution structure of the core domain (11-85) of the murine norovirus2 vpg protein Run Investigator
2	c2m4hA_	Alignment		96.7	33	**PDB header:**viral protein **Chain:** A: **PDB Molecule:**feline calicivirus vpg protein; **PDBTitle:** solution structure of the core domain (10-76) of the feline2 calicivirus vpg protein Run Investigator
3	c2mxdA_	Alignment		95.8	29	**PDB header:**viral protein **Chain:** A: **PDB Molecule:**viral protein genome-linked; **PDBTitle:** solution structure of vpg of porcine sapovirus Run Investigator
4	c1zxjB_	Alignment		42.3	33	**PDB header:**structural genomics, unknown function **Chain:** B: **PDB Molecule:**hypothetical protein mg377 homolog; **PDBTitle:** crystal structure of the hypthetical mycoplasma protein, mpn555 Run Investigator
5	c3wmmP_	Alignment		20.3	46	**PDB header:**photosynthesis **Chain:** P: **PDB Molecule:**lh1 beta polypeptide; **PDBTitle:** crystal structure of the lh1-rc complex from thermochromatium tepidum2 in c2 form Run Investigator

FFAS03



Results of FFAS03 search in PDB1018

Master-slave alignment(slide right to see more) does not show gaps in the query sequence, use ali links to display alignment between query and templates.

#	Score	Template	Links and tools	%id	First	. 10 . 20 . 30 . 40 . 50 . 60
1	-27.300	2m4h_A mol:protein length:71 Feline Calicivirus VPg protein	ali model follow..	23	3	......TYRGRG----------VALTDDEYDEWREHNASRKLDLSVEDFLMLRHRAALGA-
2	-25.000	2mxd_A mol:protein length:62 Viral protein genome-linked	ali model follow..	20	1	.................GAMAIALRDDEYDEWQDIIRDWRKEMTVQQFLDLKERALSGA-
3	-18.500	2m4g_A mol:protein length:75 Murine Norovirus VPg protein	ali model follow..	50	4	...............GVFRTRGLTDEEYDEFKKRRESRGGKYSIDDYLADREREEELLE
4	-7.870	6g45_A mol:protein length:610 Putative major capsid protein	ali model follow..	25	572	GKKKTTTRKKKGGE-GSGKKGGKKLSDKQLHDLRNL.....
5	-7.790	4kt0_K mol:protein length:128 Photosystem I reaction center subunit PsaK	ali model follow..	12	5	ARERGRSKRGIALHRLEYLKENQVFE---FGE.....

Phyre2 and FFAS03 have a consensus between the top three of both programs, albeit in a slightly different order. The top 3 consensus proteins are murine norovirus vpg protein, feline calicivirus

vpg protein, and viral protein genome-linked. The consensus fold types are 2m4h, 2mxd, and 2m4g.

2m4h: Score = -27.3 (FFAS), 96.7% confidence (Phyre2), seq ID = 23% (FFAS), 33% (Phyre2), length of overlap: 3-67 (FFAS), 7-55 (Phyre2)
2mxd: Score = -25 (FFAS), 95.8% confidence (Phyre2), seq ID = 20% (FFAS), 29% (Phyre2), length of overlap:1-62 (FFAS), 25-55 (Phyre2)
2m4g: Score = -18.5 (FFAS), 100% confidence (Phyre2), seq ID = 50% (FFAS), 50% (Phyre2), length of overlap: 4-53 (FFAS), 18-101 (Phyre2)

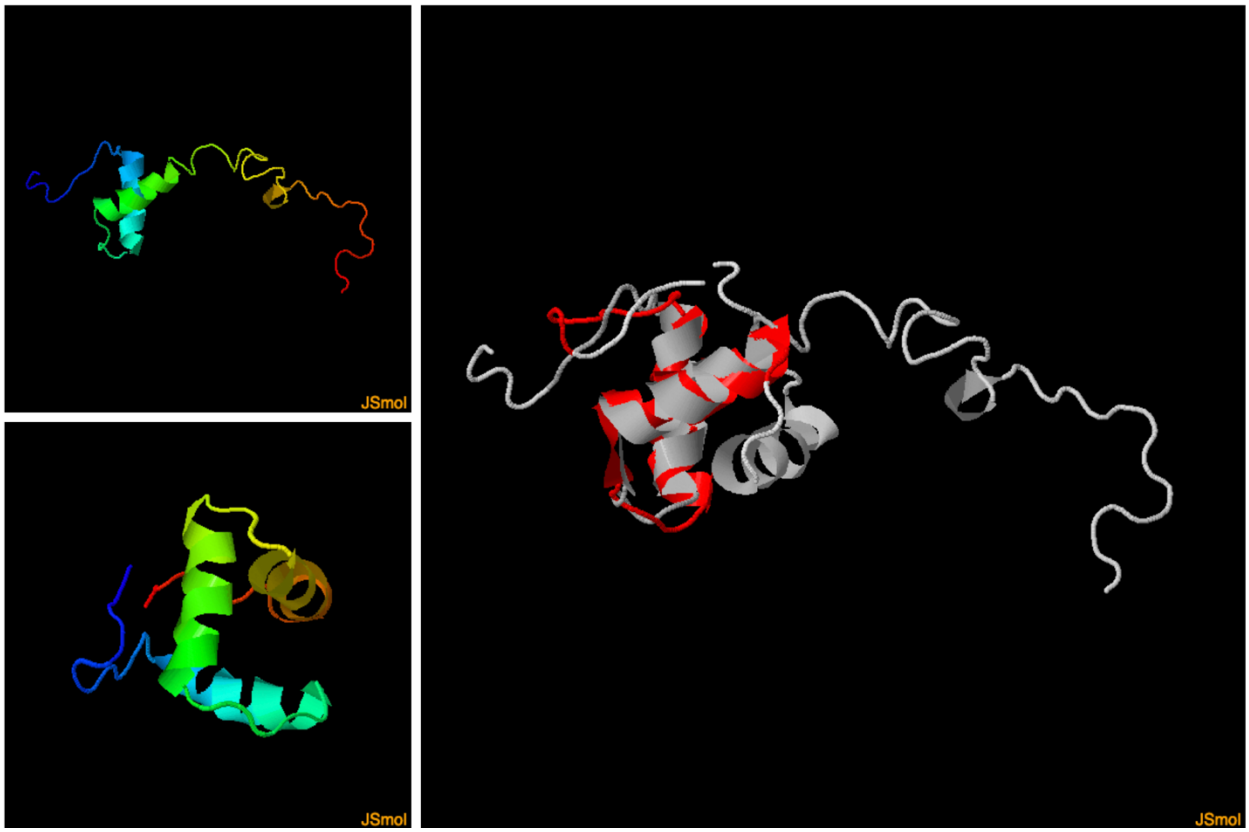
For I-TASSER

Rank	PDB Hit	Iden1	Iden2	Cov	Norm. Z-score	Download Align.	Sec
1	2m4gA	0.41	0.27	0.54	1.21	Download	
2	2mxdA	0.22	0.09	0.33	1.18	Download	
3	2m4gA	0.43	0.27	0.54	1.09	Download	
4	2m4g	0.45	0.27	0.52	5.62	Download	
5	2m4g	0.44	0.27	0.51	4.12	Download	

2m4h was missing from the top 10 hits from I-TASSER.

All three of the top consensus hits were missing from SCOP. From PDB: 2m4h is classified as viral protein, 2mxd is classified as a viral protein, and 2m4g is also classified as a viral protein.

For structural alignments, I chose to use 2m4g from I-TASSER and 2m4h from FFAS. I ran these through a flexible alignment using FATCAT and received the following output: RMSD = 2.66 and seq ID = 37.50% The top left image is 2m4hA, the bottom left is 2m4gA, and the right are the two models superimposed on each other.



Because this is a blind prediction, our alignments of these proteins give us a good starting point for further investigations. If I was interested in studying this particular protein, I would treat this information as a basis for further research on the properties and activities of the two related sequences above. The next step afterwards would be to gather empirical evidence using x-ray crystallography to determine the “actual” structure of this protein and continue investigations from there.